

VIDEO SCULPTURE: SPATIO-TEMPORAL WARPING

A Thesis

by

JEFF GROVES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2003

Major Subject: Visualization Sciences

VIDEO SCULPTURE: SPATIO-TEMPORAL WARPING

A Thesis

by

JEFF GROVES

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Donald House
(Chair of Committee)

Carol LaFayette
(Member)

John Keyser
(Member)

Phillip Tabb
(Head of Department)

December 2003

Major Subject: Visualization Sciences

ABSTRACT

Video Sculpture: Spatio-temporal Warping. (December 2003)

Jeff Groves, B.S., Portland State University

Chair of Advisory Committee: Dr. Donald House

In this thesis the concept behind our notion of video sculpture is to imagine an image sequence or movie as a three dimensional volume. We then also imagine that there is a frameset that traverses the image sequence to give us what we commonly think of as a video or movie. In the ordinary sense this frameset moves through an image sequence in a completely time-parallel linear fashion. In video sculpture, we free the frameset from these bounds so that we can sample space and time in completely unorthodox ways. We can view the when-where in previously unforeseen perspectives. Slices of the video environment can simultaneously reveal both past and future actions within a single frame.

Building on this free representation of video space-time, we then wrest the frame once more from the present constraints of topography and/or topology. The frame can bend and twist and jump and dive. The view of a fading quadratic surface cutting through two scenes makes for a beautiful curtain transition. We present a framework and an implementation for modeling the frame as it passes through the image sequence volume object.

To Beverly and Larry Charles

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Donald House, for his invaluable guidance and experience. I often relied on his ability to keep me on course. Thanks to Professor Carol Lafayette, who could always see the big picture and refused to let me fall short. Also, I would like to thank Dr. John Keyser for his knowledge and encouragement. Special thanks to Dr. Ergun Akleman, who set me on the right path early on. I am most grateful to Professor Karen Hillier for providing the environment to produce great work. Also, I would like to thank Michael Ringham, who gave so much time and attention to his VIZA652/VIZA653 students. Many thanks to Dr. Andruid Kerne, a good friend and great thinker whose help came at the perfect time.

I hold nothing but the highest regard for Michael Mistrot, who made the years in College Station an enjoyable work experience. As well, Jeff Smith was great company during the late nights while programming. Thanks to Charu Sharma Clark, Brian Clark, Lori Green, and Felice House for their friendship and feedback. Though I may try, there is no way to say thank you enough times to Miriana Ilieva, who gave me unending love and support. I would never have finished this thesis without her.

Most special thanks to my parents, Beverly and Larry Charles, who always lent a helping hand and gave such great encouragement and love. This work is a testament to their spirit and beliefs. Thanks to my sisters and their families for keeping me in their thoughts and proving the importance of family.

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
II	CONCEPT	7
III	PREVIOUS WORK	9
IV	METHODOLOGY	13
	IV.1. Image Sequence Object Representation	14
	IV.2. Creating the Sequence	18
	IV.3. Spatio-temporal Warping	18
V	IMPLEMENTATION	20
VI	RESULTS	24
VII	CONCLUSION AND FUTURE WORK	31
	REFERENCES	33
	VITA	35

LIST OF FIGURES

FIGURE	Page
1 Charles Joseph Minard's <i>Carte Figurative</i> , 1861	4
2 Series of images taken by Edison and Dickson's Kinetoscope (a), a Phenakistoscope (b).	4
3 Naum Gabo, Linear Construction No. 4	6
4 Frames rotated in frameset space (a), frames with varying topography (b) .	8
5 Early test shows a rotated frameset sample.	8
6 <i>The invisible shape of things past</i> , violin (a), Potsdamer Platz (b)	11
7 tx-transform frame space-time representation (a), rotated frame representation (b), rotated frame image result (c)	11
8 Ken Perlin's hypertexture marble vase (a) <i>Interactive video cubism</i> , time orthogonal slice (b), arbitrary time-space axis (c)	12
9 Stylized video cubes, time-varying aperture view (a), interface (b)	12
10 Example of a scene graph (a), set intersection of an image sequence \mathcal{G} and the frame set φ (b)	14
11 Object representation and reconstruction	15
12 Animation path and object location and orientation	19
13 Animation path folds back on itself in (a), path loops through space in (b)	19
14 Screen shot of the application windows	20
15 The two ways to map a voxel sample to the display space	21
16 Diagram of basic program structure	23

FIGURE		Page
17	Still no. 1 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape $f(x, y) = \frac{xy(x^2-y^2)}{(x^2+y^2)}$ (b)	26
18	Still no. 2 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape $f(x, y) = \frac{xy(x^2-y^2)}{(x^2+y^2)}$ (b)	26
19	Still no. 3 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape $f(x, y) = \frac{xy(x^2-y^2)}{(x^2+y^2)}$ (b)	27
20	Still no. 4 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape $f(x, y) = x^2$ (b)	27
21	Still no. 5 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape $f(x, y) = \frac{\sin(2x^2+3y^2)}{(x^2+y^2)}$ (b)	28
22	Still no. 8 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape $f(x, y) = x^2$ (b)	28
23	Still no. 6 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape is a raised block shape (b)	29
24	Still no. 7 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape is raised vertical bars (b)	29
25	Still no. 9 from demo video <i>journeys in time and space:spatio-temporal warping</i> (a), frame shape HSV color space value parameter (b)	30

CHAPTER I

INTRODUCTION

My total conscious search in life has been for a new seeing, a new image,
a new insight. This search not only includes the object, but inbetween places.
The dawns and the dusks.

-Louise Nevelson

This thesis will explore the foundations of a new visual form that we call video sculpture or spatio-temporal warping. In order that we may understand what video sculpture is we must look at the components, technical and historical, that define it. Webster's New World Dictionary defines sculpture as "the art of carving wood, chiseling stone, casting or welding metal, modeling clay or wax, etc. into three-dimensional representations, as statues, figures, forms, etc." [1] The key terms, "three-dimensional" and "forms," offer a linguistic basis for constructing the video sculpture framework. As the definition includes "modeling" we will use it and "sculpture" interchangeably throughout. "Video" is etymologically descended from the Latin word *videre* "see." In the vernacular we most often refer to digital video representation as a sequence of images or animation. A sequence implies an unfolding indexed by some parameter. In digital video this naturally refers to time. So, we have two-dimensional images together with some t (time) depth to create a three-dimensional sequence object. This transformation or mapping of the discrete video space to a real continuous space is the foundation for modeling video.

One hundred and seventy-three years ago, the notion of event-space was explored to explain the limitations of narrative. In 1830 historian Thomas Carlyle penned an essay on the difficult nature of writing about history.

The journal model is *IEEE Transactions on Visualization and Computer Graphics*.

The most gifted man can observe, still more can record, only the series of his own impressions; his observation, therefore... must be successive, while the things done were often simultaneous; the things done were not in a series, but in a group. It is not acted, as it is in written History: actual events are nowise so simply related to each other as parent and offspring are; every single event is the offspring not of one, but of all other events, prior or contemporaneous, and will in its turn combine with all others to give birth to new: it is an ever-living, ever-working Chaos of Being, wherein shape after shape bodies itself forth from innumerable elements. And this Chaos... is what the historian will depict, and scientifically gauge, we may say, by threading it with single lines of a few els in length! For as all Action is, by nature, to be figured as extended in breadth and in depth, as well as in length... so all Narrative is, by its very nature, of only one dimension; only travels forward towards one, or towards successive points; Narrative is linear, Action is solid.[2]

Here, Carlyle has convincingly described the limitations of the written word in accurately depicting an event while possibly anticipating some future system of representation. The physical natures of storytelling and reading as linearly time-based phenomena inhibit the multi-dimensional totality of space-time. As Carlyle describes the successive events, we are aware of the successive words, one by one, forming in the reading mind.

Much of our modern understanding of time and time-space representation is rooted at this point in the 19th century. Graphically speaking Carlyle may have been hinting at Charles Joseph Minard's unequalled 1861 chart of Napoleon's ill-fated march on Russia in 1812. Described by Tufte as "seeming to defy the pen of the historian by its brutal eloquence,"[3] this space-time graphic clearly depicts the state of Napoleons army in relation to time, temperature, and geography. We are astounded by the inconceivable attrition of 97% of his forces, yet we are, perhaps, most astonished at the simple elegance of the graphic itself. See Figure 1[3].

As Carlyle was musing on the one-dimensional nature of narrative time a new art form was stirring. Photography was at least able to capture some two-dimensional slice of time-space. This ability has not lost its glimmer. For a technological device, the essence of

photography has changed little. The major changes being the implementation of photos in rapid succession (eg. film) and the invention of CCD's to capture light digitally. These two concepts, film and digital image capture, gave birth to the digital video or digital film. It is then in the assemblage and editing of this latest medium that the idea of video sculpture was conceived.

Founded upon Peter Mark Roget's experiments with Newton's persistence of vision phenomenon, the 19th century hosted many inventions pertaining to time-based media[4]. John Paris's Thaumatrope, Joseph Plateau's Phenakistoscope, William George Horner's Zoetrope, Pierre Jules César's photographic revolver, Muybridge and Marey's chronophotography, and finally the film-strip based implementations by the Lumiere brothers in France and Edison and Dickson in the U.S.[5]. The early films were no more than single shot scenes of some one action (e.g. a kiss as in Figure 2[4][5]). Narrative crept in with Georges Melies, but editing was first used by Edwin S. Porter on *The Great Train Robbery* [6]. The concept spread quickly and developed several schools and theoretical models.

The new Russian directors founded much of the early film theory.

The film director [as compared to the theater director], on the other hand, has as his material, the finished, recorded celluloid. This material from which his final work is composed consists not of living men or real landscapes, not of real, actual stage-sets, but only of their images, recorded on separate strips that can be shortened, altered, and assembled according to his will. The elements of reality are fixed on these pieces; by combining them according to his desire, the director builds up his own 'filmic' time and 'filmic' space. He does not adapt reality, but uses it for the creation of a new reality, and the most characteristic and important aspect of this process is that, in it, laws of space and time invariable and inescapable in work with actuality become tractable and obedient. The film assembles from them a new reality proper only to itself.[7]

This idea of "filmic" space-time is instrumental to a construction of the video modeling framework.

It is evident throughout our study of sculpture that painting has always overshadowed the three-dimensional work. This is evident in the subordination of the kinesthetic to the visual in language and psychology. There is an implied superior validity to “I think” as opposed to “I feel.” There are early references to this western preeminence of the eye in the writings of Plato. He speaks of the light of reason and the eye of the soul.. From the cloudy ignorance of “the Cave”[8] to the linguistic markers insight and enlightenment there is a clear mapping between vision and understanding.

In reviewing background information on sculptural history we discovered an art critical lacuna in the theory of sculpture prior to the last half of the 20th century. At best we are left with the quotations of sculptors. “To think about sculpture as a succession of two-dimensional images would mean to think about something else, but not sculpture... sculpture is three-dimensional eo ipso.”[9] It would appear that Naum Gabo’s statement is the counter-argument to video sculpture were it not that his dynamic sculptures, such as Figure 3 below, seem often to be constructions of these “succession” ’s he derides. Indeed, Philip Rawson responds to Gabo’s work. “ No one can see a volume. All that anyone sees in a sculpture are surfaces, and from them one infers the forms of volumes that they might contain and shape, lying invisible behind or beyond them.”[10]

This surface nature of sculpture is what allows us to use video frames as our building material. In contrast to Gabo’s statement, we are specifically using successions of two-dimensional images to model our sculptural raw material. Despite Gabo’s language, he betrays his meaning with the title of his manifestation *Linear Construction No. 4*. Linear implies it is composed of a line or lines, or it is a first-degree mathematical function. By its very nature a line represents succession from one point to another. Video sculpture could be the *planar* construction analog to Gabo’s work. We need merely substitute video frames for line segments.



Fig. 3. Naum Gabo, Linear Construction No. 4

CHAPTER II

CONCEPT

The concept behind our notion of video sculpture is to imagine an image sequence or movie as a three dimensional volume. We then also imagine that there is a frameset that traverses the image sequence to give us what we commonly think of as a video or movie. In the ordinary sense this frameset moves through an image sequence in a completely time-parallel linear fashion. In video sculpture, we free the frameset from these bounds so that we can sample space-time in completely unorthodox ways. We can view the when-where in previously unforeseen perspectives. Slices of the video environment can simultaneously reveal both past and future actions within a single frame.

Building on this free representation of video space-time, we then wrest the frame once more from the present constraints of topography and/or topology. The frame can bend and twist and jump and dive. The view of a fading quadratic surface cutting through two scenes makes for a beautiful curtain transition.

One simple analogy is the loaf of bread. If we assume an image sequence exists as a basic loaf, the usual method of viewing is analogous to the usual method of slicing a loaf of sandwich bread. If we allow that the loaf can be fed into the bread slicer at any angle or orientation then the loaf will emerge not unlike our new display sequence. Then, if we alter the path of the individual blades in the slicing, we can achieve sculptural effects in the loaf. Figure 4 shows one way of visualizing this process.

Likewise, we can imagine that instead of slicing our basic box-shaped loaf of bread that we first twist, shape, push, or pull the dough, or combine two or more types of dough. We create unique morphologies which when sliced show clearly non-sandwich-like at-

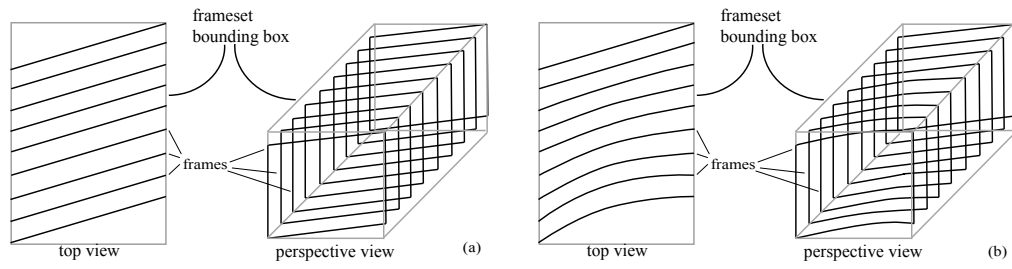


Fig. 4. Frames rotated in frameset space (a), frames with varying topography (b)

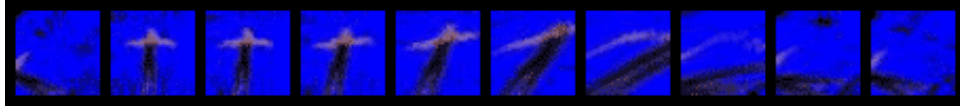


Fig. 5. Early test shows a rotated frameset sample.

tributes. Figure 5 illustrates this alternative viewing perspective.

CHAPTER III

PREVIOUS WORK

There are several existing examples of work where an animation/image sequence is treated as a solid object. It may all begin with the epipolar-plane image analysis of Bolles, Baker, and Marimont [11]. The epipolar plane is the plane that cuts through an image sequence or video cube. Whereas Bolles et al. were cutting video to analyze the movement of object lines to track motion for robotic vision the next venture into the epipolar was a completely artistic one.

In *The Invisible Shape of Things Past* [12], which recalls the overly Shakespearean translation of Proust's *À la recherche du temps perdu*, the artists use the videos as objects for sculptural display. Image sequences generate textures for rendering these sculptural pieces. They also composite these running video streams into 3D environments as well as other videos. The technique consists of extruding some surface to which the video is applied as texture. The novel approach here understood that the edge of the frame could act as texture for the outside surface. By viewing an individual frame edge on from the sides, they applied the pixel color values of these outermost edges to the surface of the evolving construction thereby creating a uniquely textured object. Figure 6 shows that while the artifice is unquestionable we cannot use the technique in an extensible manner.

A technique for slicing the time-space for practical application was developed in 1998 by M. Reinhardt [13] called tx-transform. Some fantastic results are posted and clear mappings from the frame space to the display space are shown. Figure 7[13] shows the frame space and the resultant image. The striking cubist image of all perspectives visible at once is achieved through careful planning of frame rotation. The internal camera is rotated at the same rate as the actual camera, thereby splaying the head model into a virtual skin.

There are other spatio-temporal effects, but the frame never leaves the camera-orthogonal coordinate spatial orientation. All images maintain the same ground or reference plane as the actual original footage. Likewise, we see that the planar dominance of the frame is unquestioned.

In Figure 7(a) above there are some confusing graphic elements. M. Reinhart has left some reference markers manifesting as horizontal lines. These create the impression that some horizontal slicing of the video block is taking place. This is not the case. There are no examples of such sample orientation in their implementation.

The first results of slicing a video solid in arbitrary time came from Fels and Mase in a paper from 1999 entitled *Interactive Video Cubism*[14]. They develop the (t, x, y) three-space and use a rectilinear cut-plane to slice the cube. The results, a few of which are shown in Figure 8 are texturally intriguing, with some resembling Perlin procedural marble hypertextures[15] in Figure 8[16] below. The implementation looks to be very fast, allowing for a truly interactive experience.

Klein, Sloan, Finkelstein, and Cohen take this concept a bit further in their paper *Stylized Video Cubes*[17]. They focus on the three-dimensional video as a way to do non-photorealistic rendering(NPR). The big step forward is their implementation of what they refer to as rendering solids[17]. These objects perform some set of operations on the video cube for some number of frames and section of space. The authors also explored multiple time views. The software, though, did not exploit the 3D modeling paradigm or non-planar image sampling. Manipulation still occurred on the 2D image plane. See Figure 9.

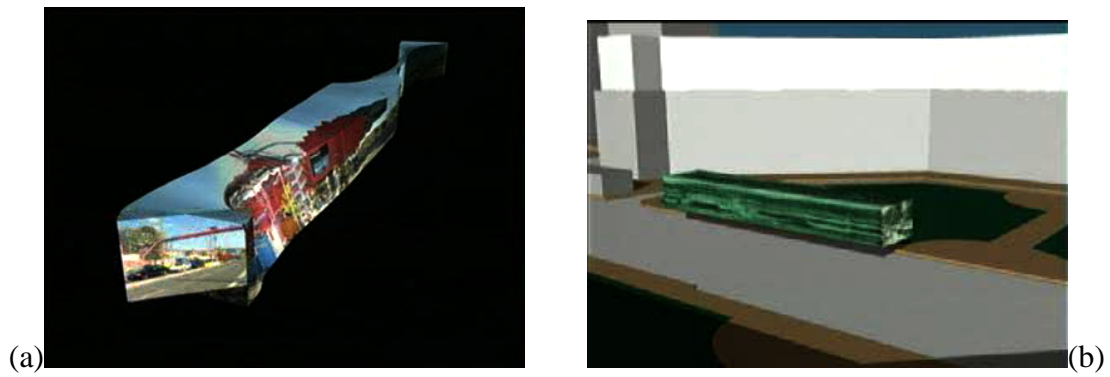


Fig. 6. *The invisible shape of things past*, violin (a), Potsdamer Platz (b)

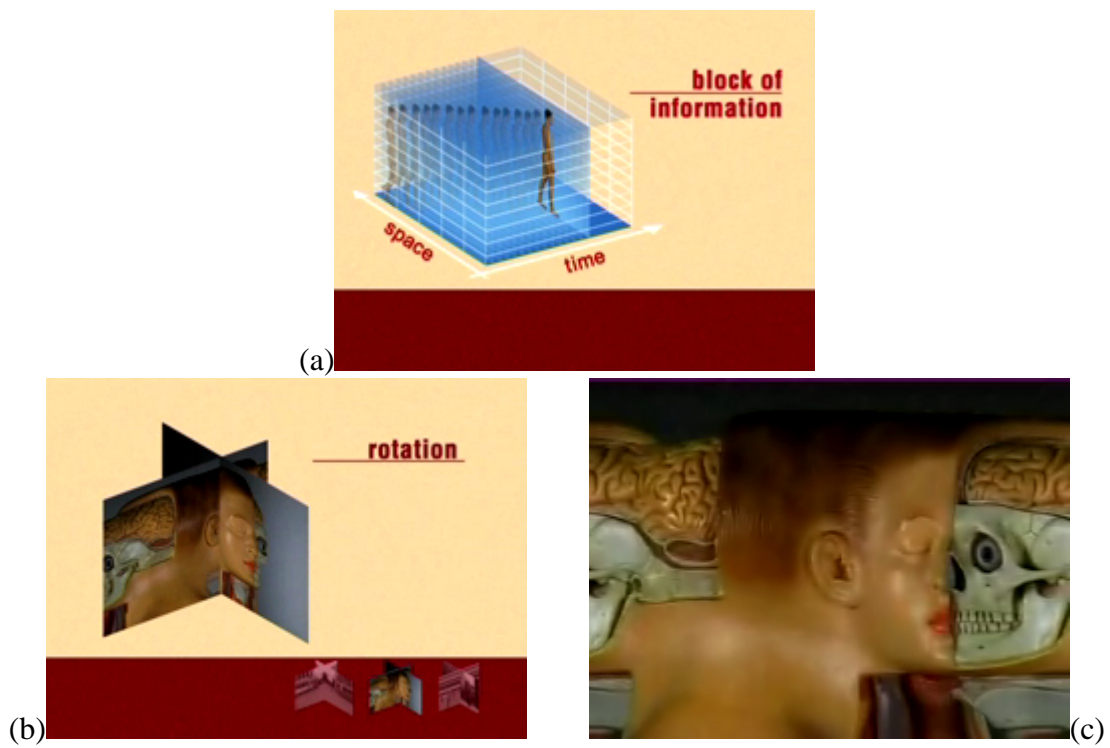


Fig. 7. tx-transform frame space-time representation (a), rotated frame representation (b), rotated frame image result (c)

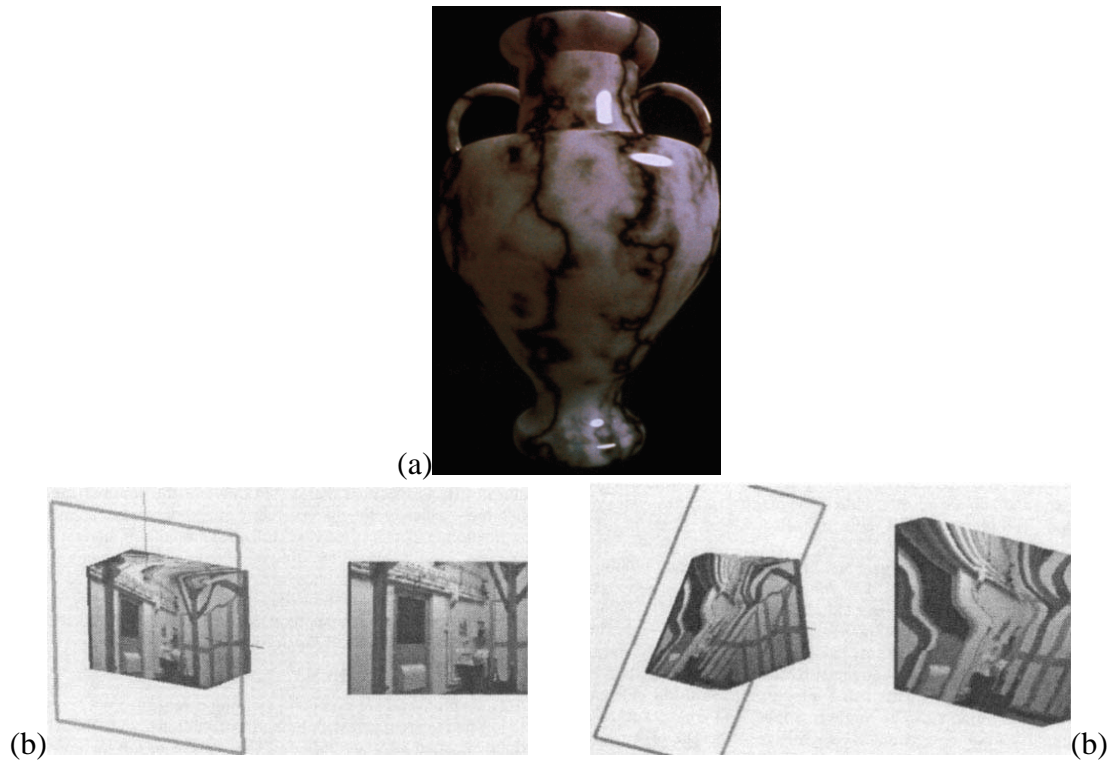


Fig. 8. Ken Perlin's hypertexture marble vase (a) *Interactive video cubism*, time orthogonal slice (b), arbitrary time-space axis (c)

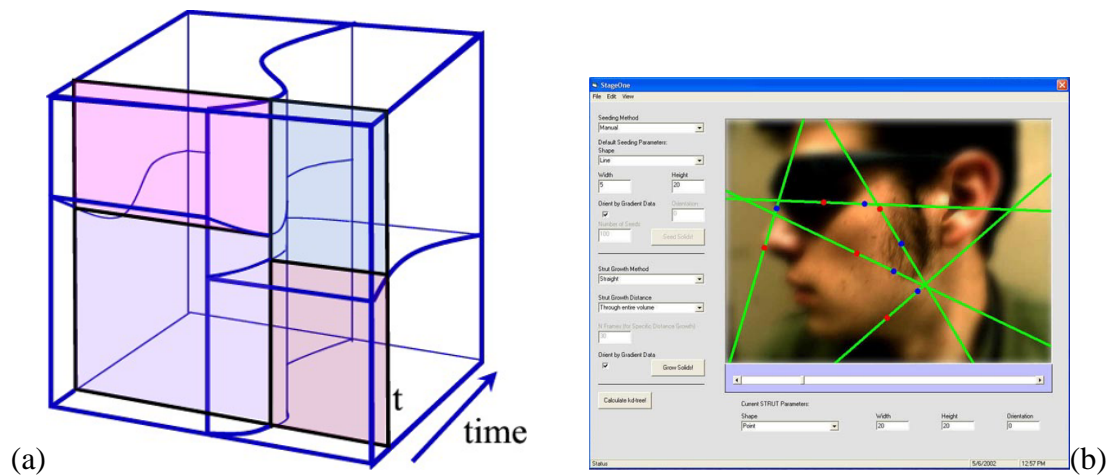


Fig. 9. Stylized video cubes, time-varying aperture view (a), interface (b)

CHAPTER IV

METHODOLOGY

Following the clearly defined graphics model established by Gomes et al in *Warping and Morphing of Graphical Objects*[18] and the earlier *The Mathematical Structure of Raster Graphics* by Fiume[19] we have constructed a mathematical-graphical model of video-space. We will introduce a new type of operator Φ - the frame-space operator. This morphism will map the real-space \mathbb{R}^3 back to display-space $\mathbb{N} \times \mathbb{N}^2$. It will also be shown that Φ is an instance of what Gomes[18] defines as a metamorphosis. By abstracting the frame object into a three-dimensional point set and differentiating it from the image sequence we will show that we can map all film, and analog videotape-based edits, transitions, and compositions, but that the representation of animation in three-space offers many more possibilities in non-linear editing. The entire set of computer modeling operations including Boolean set operations, color space functions, geometric/affine transformations, and/or user-defined expressions can be applied to a movie.

Our approach synthesizes geometric/volumetric modeling algebras with non-linear editing of digital media. As the previous examples of volumetric video dataset visualization methods use regular 3D grids, we base our image sequence representation on an equivalent structure. The videos or image sequences are imported into the application and converted into a rectilinear volume or box. The color space is represented by voxels. As the sequence is now a geometric object we can perform any number of metric transformations. The frame set object is also represented as a volume, but as a clearly discretized set of individual frames. This helps the user to relate the frame-space to the sequence space. The Φ (frame-space) operator is most easily imagined as the intersection of the frame set and the sequence set. Figure 10 shows a visual analog.

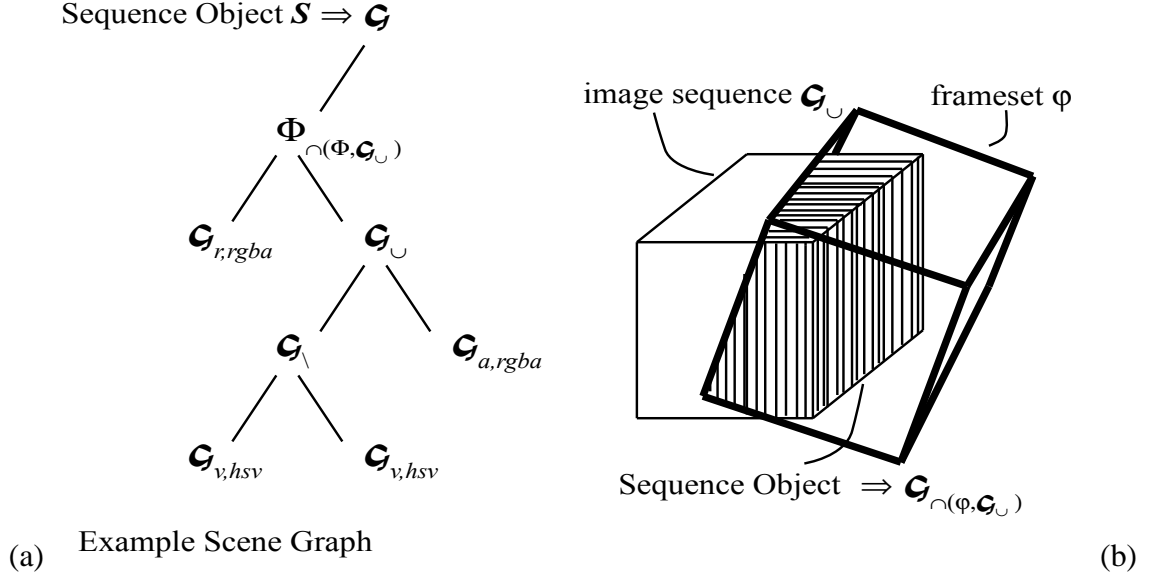


Fig. 10. Example of a scene graph (a), set intersection of an image sequence \mathcal{G} and the frame set φ (b)

IV.1. Image Sequence Object Representation

Effective description of graphical objects is required for mathematically efficacious manipulation. The issue is the cycle of reconstruction and representation shown in Figure 11. The basis for our implementation of spatio-temporal warping requires the construction of a graphical object which we build using the following definitions from Chen, Tucker, and Leu.[20]

Definition 4.0: A spatial object is a tuple $\mathbf{o} = (O, A_1, \dots, A_k)$ of scalar fields defined in \mathbf{E}^3 , including an opacity field $O : \mathbf{E}^3 \rightarrow [0, 1]$ specifying the visibility of every point in \mathbf{E}^3 and possibly other attribute fields $A_1, \dots, A_k : \mathbf{E}^3 \rightarrow [0, 1], k \geq 0$.

Definition 4.1: A scalar field $F : \mathbf{E}^3 \rightarrow [0, 1]$ is bounded if there exists a bounded set

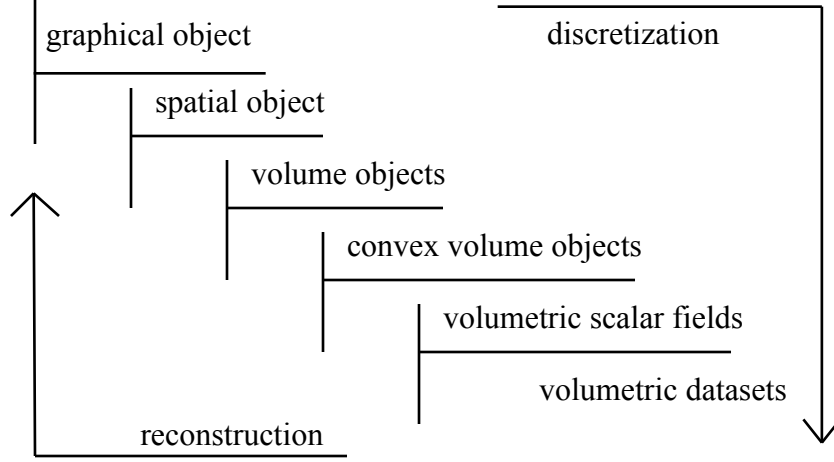


Fig. 11. Object representation and reconstruction

$X \subset \mathbf{E}^3$ such that $x \in \mathbf{E}^3 - X$ implies $F(x) = 0$.

Definition 4.2: A spatial object \mathbf{o} is a volume object if there exists a bounded set X such that each scalar field of \mathbf{o} is bounded by X .

It is necessary to explain the steps from acquiring volumetric data sets to generalizing volume objects. Again from Chen et al., given a finite set $P = p_1, p_2, \dots, p_n | p_i \in \mathbf{E}^3$ of distinct points, we call the convex hull $\mathbf{Vol}(P)$ of the point set P the volume of P , and p_1, p_2, \dots, p_n voxels. When each voxel p_i is associated with a known scalar value $v_i \in [0, 1]$ and there is an interpolation function I that uniquely determines a value at every point within $\mathbf{Vol}(P)$, a volumetric scalar field F can thus be defined in \mathbf{E}^3 as:

$$F(p) = \begin{cases} I(p, (p_1, v_1), (p_2, v_2), \dots, (p_n, v_n)) & p \in \mathbf{Vol}(P) \\ 0 & p \notin \mathbf{Vol}(P) \end{cases} \quad (4.1)$$

Definition 4.4: A convex volume object based on an interpolation method I is a vol-

ume object that consists of a finite set of volumetric scalar fields all of which are defined upon the same $\text{Vol}(P)$ by the same interpolation method.

This conceptual framework allows us to more clearly illustrate the mathematical operations that define spatio-temporal warping of video. We begin the process of reconstruction by following this chain of reasoning from the dataset to the more general graphical object. The image sequence is loaded into memory as a single block of data. This is similar in approach to the aforementioned methods. There are four scalar values $v \in [0, 1]$ for each voxel p in the dataset. These correspond to the RGBA color space members-red, green, blue, and alpha. For each member of the RGBA color model there is an associated volumetric scalar field F . In the case of an image sequence the dataset is a regular 3D grid so, for example, we might use either trilinear or Catmull-Rom spline-based interpolation methods for I in equation (4.1). The set of these four scalar fields is our convex volume object[20].

The introduction of two or more of these convex volume objects allows us to create complex constructions. CVG or constructive volume geometry provides set operations for a finely winnowed model. Using the commonly known union \cup , intersection \cap , and difference \setminus operators we can build up non-trivial structures. Figure 10 is an example of a CVG model tree created by successive CVG operations.

This constructive model concept extends to the entire scene. Based on the same paradigm as other 3D programs (e.g. games, computer-aided drafting, computer-aided modeling, and computer animation programs), the overall structure is founded upon a scene graph. Each graphical object contributes some value to the whole. In the end there is a single graphical object S (referring to sequence or scene) that is the result of the Φ operation on two objects. The first is a single graphical object comprised of all others. The other is the frame space object or frame set.

There are two ways of describing the frame space operator Φ . We begin by addressing

the set-theoretic version. If we view the frame set φ as a graphical object and more specifically as a convex volume object we will define it as a tuple $\mathbf{o} = (O, A_1, \dots, A_k)$ whose attributes correspond to the RGBA color space. Thus we have $\mathbf{o}_\varphi = (O_\varphi, R_\varphi, G_\varphi, B_\varphi)$. The operation to produce S is the intersection \cap . See Figure 10.

$$S = \mathbf{o}_G \cap \mathbf{o}_\varphi.$$

The resultant set S could, of course, be recursively operated upon by Φ to produce ever divergent graphical objects,

$$S = ((\mathbf{o}_G \cap \mathbf{o}_\varphi) \cap \mathbf{o}_\varphi) \cap \mathbf{o}_\varphi \dots$$

but we will limit this to one operation for our discussion.

We can define Φ in an algebraic fashion as well. Following Gomes et al.[18], we will create a t based parametric object or animation on the curve $c : [0, 1] \rightarrow \mathbb{R}^n$ on the parameter space where $c(0) = v_0$ and $c(1) = v_1$

$$F \circ c : \mathcal{O} \times [0, 1] \rightarrow \mathbb{R}^n.$$

For $t = 0$ we have

$$F \circ c(\mathcal{O}_1, 0) = F(\mathcal{O}_1, c(0)) = F(\mathcal{O}_1, v_0) = \mathcal{O}_1,$$

and for $t = 1$ we get

$$F \circ c(\mathcal{O}_1, 1) = F(\mathcal{O}_1, c(1)) = F(\mathcal{O}_1, v_1) = \mathcal{O}_2.$$

So, here we have a graphical object \mathcal{O}_1 that transforms into \mathcal{O}_2 as the parameter moves from 0 to 1. The curve c is called an *animation path*. Figure 12 below shows an example of an animation path in the xyz parameter space. The typical animation path for spatio-temporal warping of an image sequence requires at least 10 parameters

$$x, y, z, \varphi, \psi, \theta, r, g, b, a.$$

The first six define a position and orientation in \mathbb{E}^3 while the others give the location in the RGBA color space.

IV.2. Creating the Sequence

While there are at least 10 parameters required to represent the object attribute space, the actual number of parameters is much higher. We have created a frameset model that allows the user to define each frame sampled at t by its orientation and location in frameset coordinate space which is then defined in world coordinate space. See Figure 4 above.

In addition, we can alter the topography, even topology, of the sampled frame. There exist infinitely many implicit functions and color space functions to change the frame shape. We can also choose animation paths for our parameter t (e.g. linear, circular, helical...). Figure 13 shows some examples. In this case the path affects the relative weighting of the topographic equation. In Figure 13(a) we move from $c(0) = \mathcal{O}_{\Phi_1}$ to $c(0.5) = \mathcal{O}_{\Phi_2}$ to $c(1) = \mathcal{O}_{\Phi_1}$. In Figure 13(b) we begin with \mathcal{O}_{Φ_1} and finish on \mathcal{O}_{Φ_2} . At present for all cases $c(0)$ corresponds with the first frame in the frameset and $c(1)$ with the last frame.

IV.3. Spatio-temporal Warping

The introduction of the Φ operator is used to facilitate the creation of spatio-temporal warps. It should be obvious that this term refers to a warping in space(xy) and in time(t). most of us are familiar with warps of the two dimensional plane. Image manipulation and creation programs such as Adobe's Photoshop or Macromedia's Freehand allow us to apply any number of defined warps(filters) or transformations.

There are also three dimensional warps available in the current modeling packages like Alias' Maya, NewTek's LightWave, or 3D StudioMax. As these applications are also animation programs, they have the ability to warp parametrically by t through keyframing and/or tweening. This addition of t to the transformation gives us the temporal in spatio-temporal.

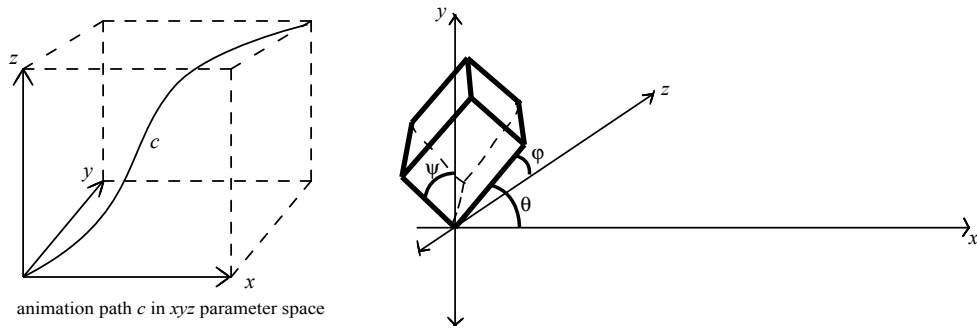


Fig. 12. Animation path and object location and orientation

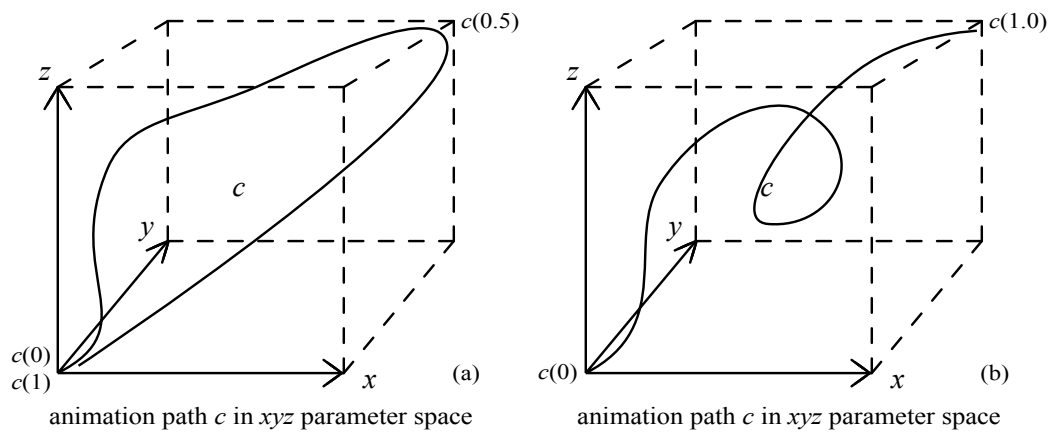


Fig. 13. Animation path folds back on itself in (a), path loops through space in (b)

CHAPTER V

IMPLEMENTATION

In order to show valid results of the proposed graphical model, we have created video sequences by applying this structure. We modeled the video-space and the frame-space in a computer application using C++ and OpenGL. The program uses what is a fairly common 3D computer modeling interface shown in Figure 14. The user has control over the particular view of the video space including perspective and orthogonal viewpoints. There is also a window that displays the more common 2D sequenced animation. Transformation in the model-space \mathbb{R}^3 is shown in the display-space $\mathbb{N} \times \mathbb{N}^2$.

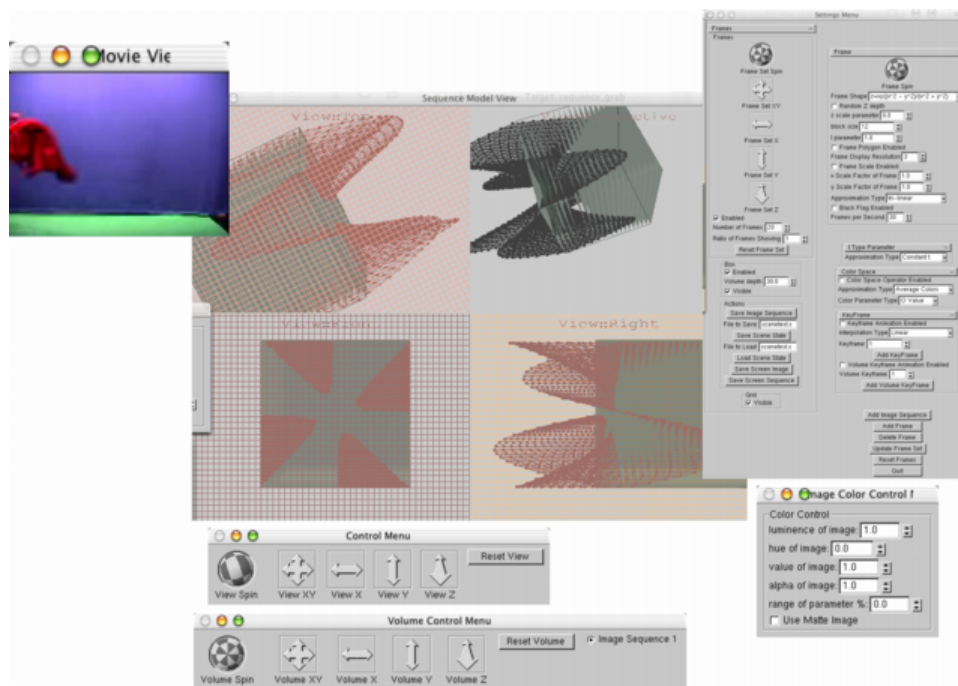


Fig. 14. Screen shot of the application windows

Our implementation of spatio-temporal warping involves the conversion of a three di-

mensional spatial object (x, y, z) into a discrete three dimensional spatio-temporal (x, y, t) object $f(x, y, z) : \mathbb{R}^3 \rightarrow \mathbb{N} \times \mathbb{N}^2$. This allows us to slice through a real-space \mathbb{R}^3 convex volume object and sample the intersecting voxels. A change in the z in our modeling space translates to a change in t in display space. They become homeomorphic, i.e. their functions map identically. There are two ways we can reference the sample and map it to the display. In this example we are using a method where the sample distance varies along the function path $f(x, y)$. One could also say that it is fixed in the xy plane. See Figure 15 below.

The unique aspect of our method gives us almost limitless control of the warping

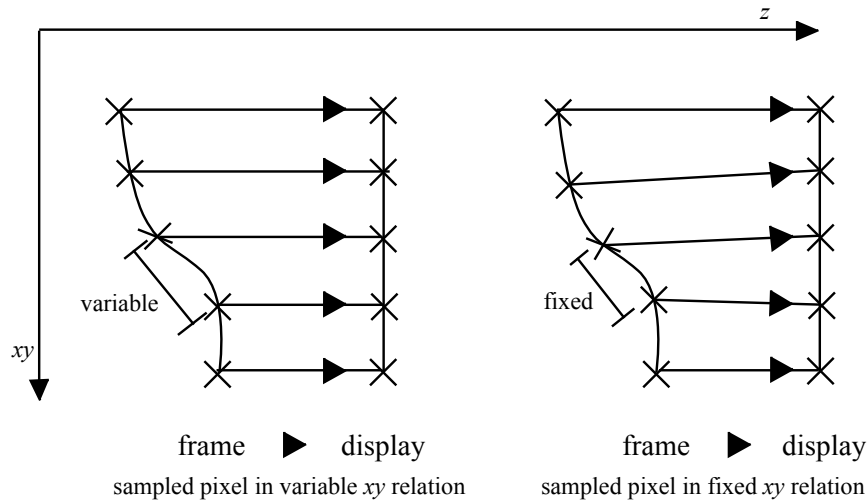


Fig. 15. The two ways to map a voxel sample to the display space

process. The linking of different attributes to the frame topography presents us with multivarious possibilites. We can set the frame to alter its shape in response to changing hue, saturation, value, luminance, or RGBA color attributes. The use of animated mattes linked

to the frame allows us very creative titling and motion graphic opportunities. This ability to vision past, present and future in a single instant through geometric means affords us the tools to describe all temporally-based media and their effects in graphical-geometrical terms.

The topography or topology of the sampled frame are generally malleable. The participant is given a list of over a dozen implicit functions and several color space functions to change the frame shape. NURBS modelling of the frame is currently alpha stage. The user may choose an animation path shape to affect the parameter t (e.g. linear, circular, helical...). This t is presently a weighting factor for the frame shape equation. See Figure 13.

The application was designed in C++ to be as extensible as possible. There were obvious object choices available which followed naturally from the theoretical framework. In brief, we developed a few major objects including a `FrameSetObject`, an `ImageSequenceVolumeObject`, an `ImageObject`, and a `FrameObject`. These primary types form the core of the spatio-temporal warping algorithm. Other supporting objects include a `ViewportObject`, `TransformObject`, `SceneStateObject`, `KeyFrameObject`, `ColorSpaceObject`, `QuaternionObject`, `MatrixObject`, `VectorObject`, and `PathAndFileObject`.

Refer to Figure 16 for a diagrammatic outline of the program structure.

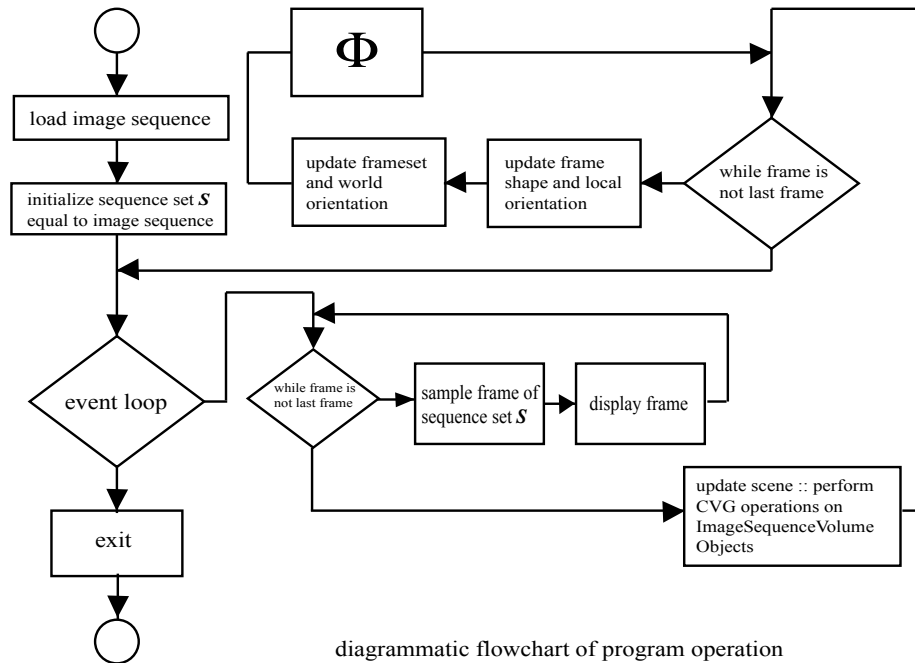


Fig. 16. Diagram of basic program structure

While rendering time on reasonably small datasets is insignificant, there are still needed improvements in optimization in order to ensure a response-time quick enough to keep an interactive environment.

CHAPTER VI

RESULTS

It is the case, then, that the same is true of film as has long been proven with every other art form. To be bound to nature is a restriction.” - Hans Rohr [21]

Many exciting images emerged from this experiment. The following figures conform to a display structure where each warped result is paired with its corresponding topography. Figure 17 shows the result of warping the transition of two shots with a frame topography defined by $f(x, y, t) = \frac{txy(x^2-y^2)}{(x^2+y^2)}$ as t varies linearly from 0 to 1 to 0 over the range of the two shots. This particular frame shape proved to be one of the most visually intriguing morphologies for relatively stable camera views. Figures 17[22] and 18[22] show this type of sequence. We can see that the result of such a static shot gives a funhouse mirror effect that warps in time and space. Figure 19[22] shows that when there is too much lateral motion in the camera view we can no longer discern clear identifiable objects within the image.

One goal of this experiment was the specification of conventional three dimensional (*height, width, time*) video and/or film transitions through four dimensional (x, y, z, t) frame set metamorphoses. One example would be wipes.

A wipe at its most basic is an operation that acts on the edit or transition of two shots. At present most applications render transitions in the color space. This means it is a compositing or blending operation. The pixels in each image are blended or not blended over some time t that defines the length of the transition.

Our method renders transition in the affine space. This allows us to represent complex and, of course, simple transitional effects. Figure 20[22] shows a simple center verti-

cal wipe that is created by a frame shape conforming to $f(x, y, t) = x^2t$ where t varies from 0 to 1 to 0 over the transition range. Another version of this type of wipe could use $f(x, y, t) = |x|t$. It is obvious to point out that horizontal versions of these wipes merely require changing the x for the y . The effect can be further augmented by applying blending at the edge conditions.

An example of a more complex member of the circular wipe family is shown in Figure 21[22]. Here we can see a ripple effect that emulates a stone thrown in a pond. More conventional circular wipes could be constructed by implementing a frame shaped like a hollow cone or semi-sphere.

These basic transitions are useful, but there are many more effects available through the use of spatio-temporal warping. Wipes are now one instance of a larger, more generic framework for image sequence manipulation. By modeling the frame itself, we are afforded a simpler mapping of effect to result. It makes sense to see the quadratic frame pass through the sequence while the shots transition one to the other as a center wipe. See Figure 22[22]. Likewise, Figures 23[22] and 24[22] present unique warps/transitions that offer much clearer analogs in the geometric space than in the color space. We can *know* what we are about to see rendered from the frame morphology.

The frame shape is not limited to our arbitrary modeling. Figure 25[22] shows the result and topology of a frame modeled on the corresponding colorspace of the image sequence. This gives us the ability to specify compositing relationships in the affine or geometric space. We can alter our frame using any set of images. In this image we have placed a grayscale matte of some text and modeled the frame based on the matte's luminance. The luminance value scales the frame in the frame local z direction. The visual effect is that the title emerges subtly from the image. Depending on some global scale factor for the set we can move the frame deep into the sequence to reveal far future frames or past action. This linkage of image color space to frame topology leads to unique phenomeno-

logical aspects. The revealed text or title exists as a time-displaced sectional version of the original sequence. Unlike chroma-keying and colorspace compositional technique where the operation is analogous to pasting atop or cutting a hole, this operation uses the matte image to define a region that warps in relation to itself. The past, present, and future exist simultaneously as inherent conditions of the technique.

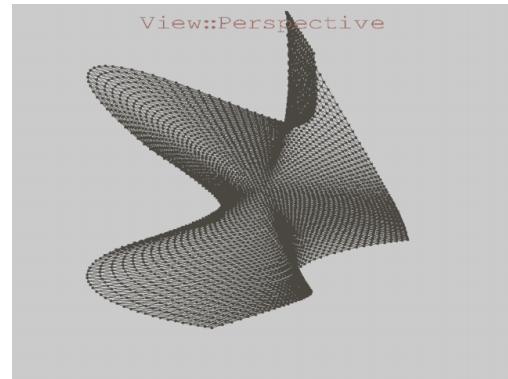


Fig. 17. Still no. 1 from demo video *journeys in time and space:spatio-temporal warping*
 (a), frame shape $f(x, y) = \frac{xy(x^2 - y^2)}{(x^2 + y^2)}$ (b)

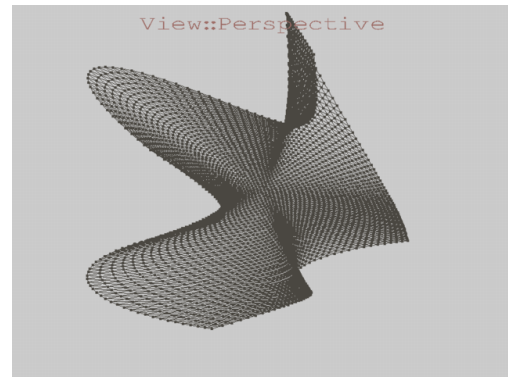


Fig. 18. Still no. 2 from demo video *journeys in time and space:spatio-temporal warping*
 (a), frame shape $f(x, y) = \frac{xy(x^2 - y^2)}{(x^2 + y^2)}$ (b)

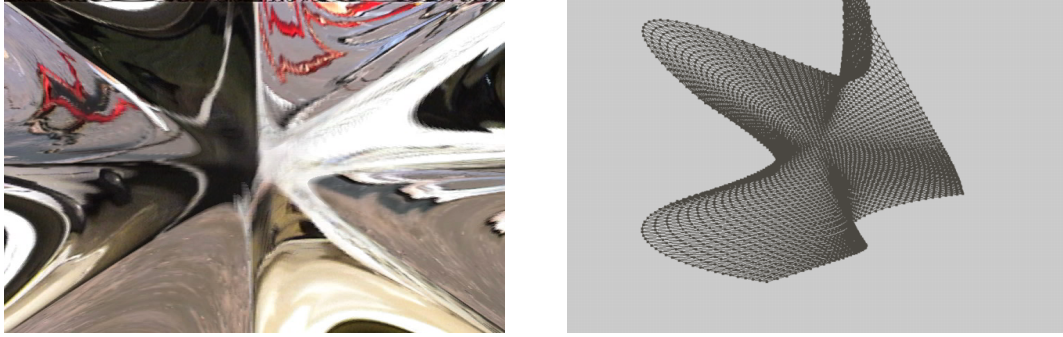


Fig. 19. Still no. 3 from demo video *journeys in time and space: spatio-temporal warping*
 (a), frame shape $f(x, y) = \frac{xy(x^2 - y^2)}{(x^2 + y^2)}$ (b)

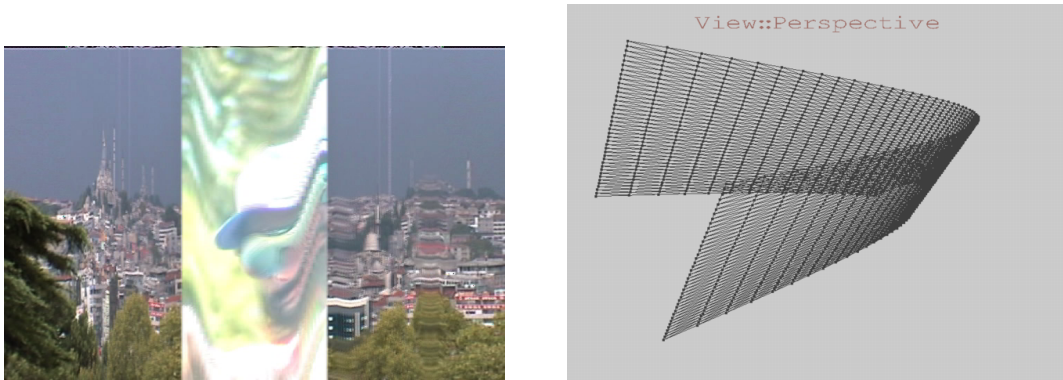


Fig. 20. Still no. 4 from demo video *journeys in time and space: spatio-temporal warping*
 (a), frame shape $f(x, y) = x^2$ (b)

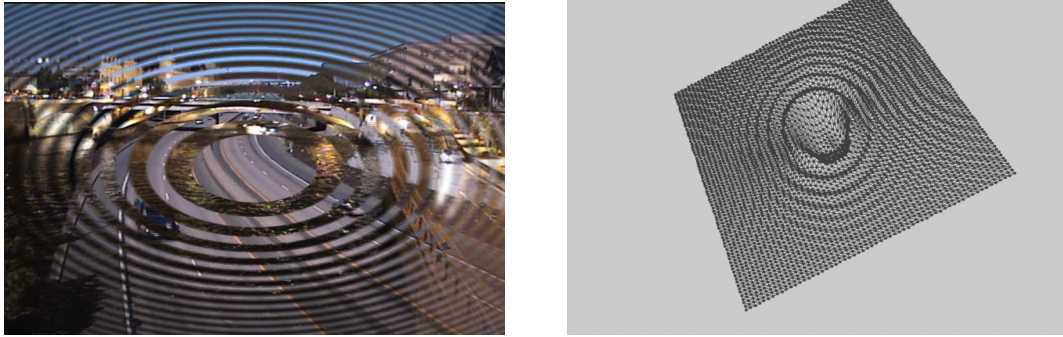


Fig. 21. Still no. 5 from demo video *journeys in time and space: spatio-temporal warping*
 (a), frame shape $f(x, y) = \frac{\sin(2x^2 + 3y^2)}{(x^2 + y^2)}$ (b)

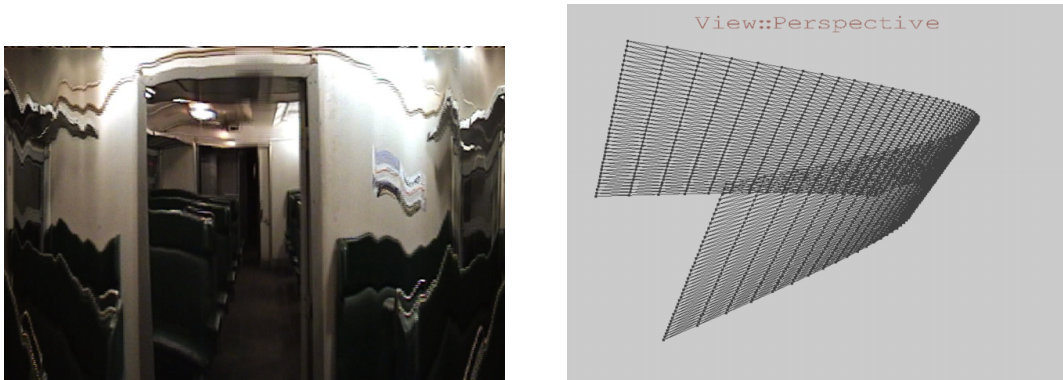


Fig. 22. Still no. 8 from demo video *journeys in time and space: spatio-temporal warping*
 (a), frame shape $f(x, y) = x^2$ (b)

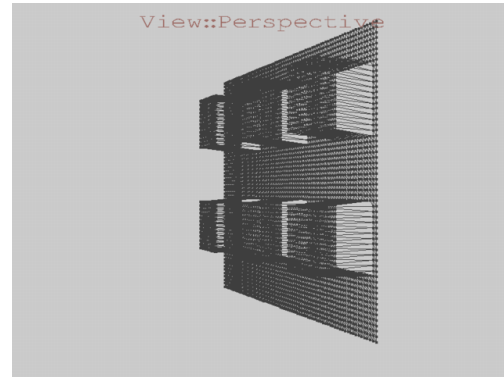


Fig. 23. Still no. 6 from demo video *journeys in time and space:spatio-temporal warping*
(a), frame shape is a raised block shape (b)

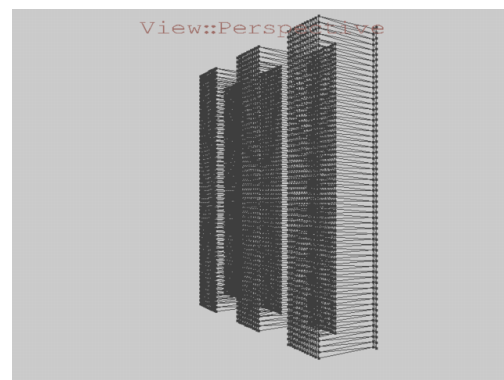


Fig. 24. Still no. 7 from demo video *journeys in time and space:spatio-temporal warping*
(a), frame shape is raised vertical bars (b)



Fig. 25. Still no. 9 from demo video *journeys in time and space:spatio-temporal warping*
(a), frame shape HSV color space value parameter (b)

CHAPTER VII

CONCLUSION AND FUTURE WORK

The proliferation of consumer digital video as well as non-linear editing software packages invites a new editing model. The current film-based paradigm is over one hundred years old. Compositing and editing are still performed almost exclusively as 2D palette operations. This framework appears limiting with respect to new effects architectures present in films like the Matrix. Use of simultaneous multi-perspective camera work should beg a comparable post-production space. The seed of such a system is the emphasis of this thesis. Future enhancements would include a scene graph/object graph modeling architecture. Graphical objects or information objects would exist as terminal nodes connected to a non-terminal operator node. Any parameter of the graphical object could map to any other objects parameter. This system would extend the constructive volume geometry model.

We find that the liberation of the frame from its current linear path and rectilinear shape gives us a unique manipulative framework for inventing the new digital editing vernacular. Spatio-temporal warping is a subset of a larger video-modeling paradigm. This work is the first step in the creation of a complete mathematical-graphical description of dimensional media space. The use of this technique is not limited to image sequences. It can operate upon audio, streams, ascii, or any digital input.

Additional features might include isosurface construction of segmented image features. This would enable parametric modeling of video or image sequences. Other uses of this ability would be object tracking, and feature extraction. Uses of such a video system could allow for easy metamorphosis of video objects, stitching of multiple camera views, more fluid temporal compression/expansion, enhanced compositional and compositing capabilities. Far from arbitrary cuts in time-space we can now define warps in relation to

some action in affine space or color space. With improved object tracking we can link objects in the sequence to actions in the frame space. Our warps are motivated by the image sequence inhabitants. While there are clear abstract aesthetic uses in a non-orthogonal time line, the true power of the system lies in understanding time in the pursuit of the action.

REFERENCES

- [1] *Webster's New World Dictionary of the American Language*. New York: Prentice Hall, 1986.
- [2] T. Carlyle, "On history," in *The Varieties of History: From Voltaire to the Present* (F. Stern, ed.), New York: Vintage Books, 1972.
- [3] E. R. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.
- [4] E. Leslie, *Hollywood Flatlands : Animation, Critical Theory and the Avant-Garde*. London: Verso, 2002.
- [5] E. Toulet, *Birth of the Motion Picture*. New York: Harry N. Abrams, Inc, 1995.
- [6] K. Dancyger, *The Technnique of Film and Video Editing: History, Theory, and Practice*. Boston, MA: Focal Press, 2002.
- [7] V. I. Pudovkin, *Film Technique and Film Acting*. New York: Grove Press, 1960.
- [8] Plato, "The Republic," in *The Collected Dialogues of Plato, Including the Letters* (E. Hamilton and H. Cairns, eds.), vol. xxv of *Bollingen*, New York: Pantheon Books, 1961.
- [9] F. D. Martin, *Sculpture and Enlivened Space: Aesthetics and History*. Lexington, KY: University Press of Kentucky, 1981.
- [10] P. Rawson, *Sculpture*. Philadelphia, PA: University of Pennsylvania Press, 1997.
- [11] R. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar plane image analysis: An approach to determining structure from motion," *Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.

- [12] J. Sauter and D. Lusebrink, “The invisible shape of things past.” <http://www.artcom.de>, 1997.
- [13] M. Reinhart, “tx-transform.” <http://www.tx-transform.com>, 1998.
- [14] S. Fels and K. Mase, “Interactive video cubism,” in *Proceedings of the 1999 Workshop on New Paradigms in Information Visualization and Manipulation in Conjunction with the Eighth ACM International Conference on Information and Knowledge Management*, pp. 78–82, 1999.
- [15] K. Perlin and E. Hoffert, “Hypertexture,” in *Proceedings of the 1989 ACM SIGGRAPH Conference*, vol. 23, pp. 253–262, 1989.
- [16] K. Perlin, “marble hypertexture.” <http://mrl.nyu.edu/perlin/doc/vase.html>, 1998.
- [17] A. W. Klein, P. Sloan, A. Finklestein, and M. Cohen, “Stylized video cubes,” in *Proceedings of the ACM SIGGRAPH Symposium on Computer Animation*, pp. 15–22, 2002.
- [18] J. Gomes, L. Darsa, B. Costa, and L. Velho, *Warping and Morphing of Graphical Objects*. San Francisco, CA: Morgan Kaufman Publishers, 1999.
- [19] E. L. Fiume, *The Mathematical Structure of Raster Graphics*. San Diego, CA: Academic Press, Inc, 1989.
- [20] M. Chen, J. V. Tucker, and A. Leu, “Constructive representations of volumetric environments,” in *Volume Graphics* (M. Chen, A. E. Kaufman, and R. Yagel, eds.), London: Springer-Verlag, 2000.
- [21] H. Rohr, “Film gegner von heute-filmfreunde von morgen,” 1929, quoted in E. Leslie, *Hollywood Flatlands : Animation, Critical Theory and the Avant-Garde*. London: Verso, 2002.
- [22] J. Groves, *Journeys in Space and Time*. digital video, 2003.

VITA

Jeff Groves

1402 Kirkwood
Unit A
Austin, TX 78722
jgroves@viz.tamu.edu

Education

M.S. in Visualization Sciences	Texas A&M University, 12/03
B.S. in Architecture	Portland State University, 6/00

Research Interests

Film and Video Production
Non-Photorealistic Rendering
Film and Video Effects
Special Effects
Physically-Based Modeling
Multimedia Production

Employment

Texas A&M University, College Station, TX	Graduate Assistant/Trainer, 9/01 - 12/03
Bainbridge Design, Portland, OR	Materials Librarian, 5/99 - 5/00
Blowout Entertainment, Portland, OR	Assistant Purchasing Director, 8/97 - 5/98
Borders Books & Music, Portland, OR	Store Trainer/Periodicals Czar, 7/94 - 8/97
United States of America,	Troubadour Musician, 9/93 - 6/94
Giant Records/Warner Bros., L.A., CA	Recording Artist, 10/90 - 3/93

Honors

n-Space Art Gallery, SIGGRAPH 2001
Graduate Merit Fellowship, Texas A&M University, 2000
Excellence in Drawing Award, Portland State University, 1999
RIAA Platinum Certification, 1999
National Merit Scholarship, 1985